

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平5-122366

(43)公開日 平成5年(1993)5月18日

(51)Int.Cl.⁵

H 0 4 M 3/50

識別記号

庁内整理番号

F I

技術表示箇所

A

審査請求 未請求 請求項の数1(全 9 頁)

(21)出願番号 特願平3-281540

(22)出願日 平成3年(1991)10月28日

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 小山内 博行

東京都青梅市末広町2丁目9番地 株式会

社東芝青梅工場内

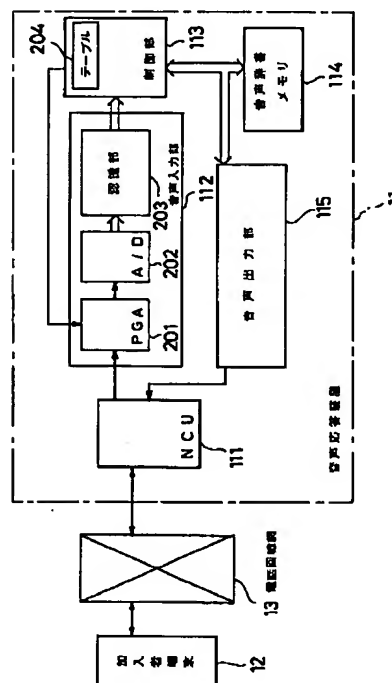
(74)代理人 弁理士 鈴江 武彦

(54)【発明の名称】 音声応答装置

(57)【要約】

【目的】入力音声信号の増幅率の制御を適切に行えるようにし、音声信号の認識率の向上を図る。

【構成】音声応答装置11に入力された音声信号の各単語種は、補正值テーブル204内の対応する補正值によって補正され、最も発声レベルの高い単語(「サン」、「ハイ」)が入力された場合と同等の音声レベルに変換される。そして、その補正結果に基づいてプログラマブル増幅器(PGA)201の増幅率が制御される。このため、各単語種毎にプログラマブル増幅器(PGA)201の増幅率の調整度合いを変更できるようになり、単語種による音声信号レベルの相違にとらわれず、常に入力した音声信号が最適なレベルになる様な増幅率の調整を行うことができる。



BEST AVAILABLE COPY

【特許請求の範囲】

【請求項1】 発信者から送信される入力音声信号に
応答して所定の音声メッセージを返送する音声応答装置に
おいて、

前記入力音声信号を増幅して入力する入力手段と、
この入力手段によって入力された音声信号を認識する音
声認識手段と、

この音声認識手段で認識可能な各単語種の音声信号レ
ベルを前記単語内で最も発声レベルの高い単語種が入力さ
れた場合と同等の音声レベルに変換するための補正值が
前記各単語毎に定義された補正值テーブルと、

前記音声認識手段によって認識された音声信号の単語種
に対応した前記補正值を用いて前記音声信号レベルの値
を補正する補正手段と、

この補正手段によって補正された補正結果と所定の基準
レベルとの比較し、その比較結果に基づいて前記入力手
段の増幅率を制御する制御手段とを具備することを特徴
とする音声応答装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】この発明は音声応答装置に関し、
特に発信者から送信される入力音声信号に
応答して所定の音声メッセージを返送する音声応答装置に関する。

【0002】

【従来の技術】一般に、発信者と音声応答装置間の音声
通信は、電話回線網を介して次のような手順で行われて
いる。まず、発信者が電話回線を介して音声応答装置を
呼出し、その呼出された音声応答装置は発信者に
応答を返す。その後、音声応答装置は、発信者から送信される
音声信号を認識し、その認識結果に対応した音声メッセ
ージを発信者に返信する。

【0003】このような電話回線を利用した音声通信に
おいては、受話器の性能や、交換器の性能、さらには伝
送路の抵抗その他の影響によって、音声信号の減衰やノ
イズの混入が発生し易い。

【0004】特に、音声応答装置に入力される発信者か
らの音声信号のレベルにばらつきがある場合には、高レ
ベルの音声信号と低レベルの音声信号が混在され、これ
によって音声認識の認識率が著しく低下される。

【0005】例えば、入力音声信号のレベルが大き過ぎ
る場合には、音声応答装置のA-D変換器で量子化する
際にそのA-D変換器の入力信号レベルの最大値を越え
てしまい、また、逆に入力音声信号のレベルが小さ過ぎ
る場合には、A-D変換器が最適に量子化しきれないとい
う不具合が生じる。このようなA/D変換器による
変換特性の悪化は、音声認識の認識率を低下させる大き
な要因となる。

【0006】そこで、従来では、音声信号のレベルの最
大値が、ちょうど、A-D変換器の入力信号レベルの最
大値になる様に、以下の方法を用いて入力信号のゲイン

調整を行っている。

【0007】一つは、A-D変換器の前に増幅器を置
き、その増幅率を可変抵抗器等で調整する方法である。
しかしながら、この方法では、一度増幅率を設定した後
はそれ以降容易にその増幅率を調整する事ができないの
で、発信者の声の大小や、発信者が電話を架けた場所の
違い等に音声信号レベルにバラツキが生じると、その音
声信号レベルの変化に十分に対応することができないとい
う不具合が生じる。

【0008】もう一つは、A-D変換器の前にプログラ
マブル増幅器(PGA)を置き、入力した音声信号のレ
ベルを検出し、そのレベルが低かったら増幅率を上げる
様に、高かったら増幅率を下げる様にプログラマブルに
増幅器の設定を変える方法である。

【0009】この方法を使用すると、可変抵抗を使用し
た方法に比べ、入力音声信号のレベルを容易に調整する
事ができる。しかしながら、入力音声信号のレベルは認
識する単語種によって異なるため、音声信号の入力レ
ベルによって一律に増幅率を変えると例えば次のような不
具合が発生する。

【0010】すなわち、一般に認識単語で数字の“イ
チ”は数字の“サン”より音声信号レベルが最大値で1
/2以下と小さい為、“イチ”を受けてそのレベルに合
わせて増幅率を上げると、その後、“サン”等の音声信
号レベルの大きい単語を受けた時、A-D変換器の入力
信号の最大値を越えてしまい認識率を下げる等の不具合
が生じることがある。

【0011】このように、従来では、発信者の声の大
小、及び電話回線での音声信号の減衰等を入力時の増幅
率の可変で補ってはいるが、その増幅率の制御を適切に
行うことができず、十分な認識率を得ることができな
かった。

【0012】

【発明が解決しようとする課題】従来の音声応答装置で
は、入力音声信号の増幅率の制御を適切に行うことがで
きず、これによって音声認識の認識率が低下されるとい
う欠点があった。

【0013】この発明はこのような点に鑑みてなされた
もので、入力音声信号のレベル不足、又はレベル過多に
よる誤認識が生じないように入力音声信号の増幅率の制
御を適切に行えるようにして、十分に認識率の高い音声
認識を行うことができる音声応答装置を提供することを
目的とする。

【0014】

【課題を解決するための手段および作用】この発明は、
発信者から送信される入力音声信号に
応答して所定の音声メッセージを返送する音声応答装置において、前記入
力音声信号を増幅して入力する入力手段と、この入力手
段によって入力された音声信号を認識する音声認識手段
と、この音声認識手段で認識可能な各単語種の音声信号

レベルを前記単語内で最も発声レベルの高い単語種が入力された場合と同等の音声レベルに変換するための補正值が前記各単語毎に定義された補正值テーブルと、前記音声認識手段によって認識された音声信号の単語種に対応した前記補正值を用いて前記音声信号レベルの値を補正する補正手段と、この補正手段によって補正された補正結果と所定の基準レベルとの比較し、その比較結果に基づいて前記入力手段の増幅率を制御する制御手段とを具備することを特徴とする。

【0015】この音声応答装置においては、入力音声信号の各単語種は補正值テーブル内の対応する補正值によって補正され、最も発声レベルの高い単語が入力された場合と同等の音声レベルに変換される。そして、その補正結果は所定の基準レベルと比較され、その比較結果に基づいて入力手段の増幅率が制御される。このため、各単語種毎に増幅率の調整度合いを変更できるようになり、単語種による音声信号レベルの相違にとらわれず、常に入力した音声信号が最適なレベルになる様な増幅率の調整を行うことができる。

【0016】したがって、発声レベルの低い単語種の後に発声レベルの高い単語種が入力された場合でも増幅率を上げ過ぎるといった不具合を防止でき、入力音声信号のレベル不足、又はレベル過多による誤認識を防止できるようになり、音声認識の認識率を十分に向上させる事が可能となる。

【0017】

【実施例】以下、図面を参照してこの発明の実施例を説明する。図1にはこの発明の一実施例に係わる音声応答装置のシステム構成が示されている。

【0018】この音声応答装置11は、加入者端末12の電話機を使用する発信者と電話回線網13を介して音声通信を交わすために、ネットワーク制御ユニット(NCU)111、音声入力部112、制御部113、音声辞書メモリ114、および音声出力部115を備えている。

【0019】ネットワーク制御ユニット111は、電話回線網13と音声応答装置11を回線接続するための網制御装置であり、このネットワーク制御ユニット111によって発信者と音声応答装置11間における音声信号の授受が行われる。

【0020】音声入力部112は、ネットワーク制御ユニット111によって受信された音声信号の入力および認識処理を行う。この認識処理では、発信者から送信される「イチ」、「ニ」、「サン」等の数値や、「ハイ」、「イイエ」等の応答文を初めとする、各種単語が認識される。この音声入力部112には、プログラマブル増幅器(PGA)201、A-D変換器202、および音声認識部203が設けられている。

【0021】プログラマブル増幅器(PGA)201は、入力した音声信号レベルを増幅してA-D変換器2

02に出力するためのものであり、制御部113の制御に応じてその増幅率が可変設定されるように構成されている。A-D変換器202は、プログラマブル増幅器(PGA)201からの音声出力を音声認識部203で処理できる様にデジタル化する。音声認識部203は、前述したような各種単語の認識を行う。

【0022】制御部113は、音声応答装置11全体の制御を行うものであり、音声認識部203からの認識結果に基づいて返送すべき音声データを音声辞書メモリ114から取り出し、それを音声出力部115に供給する。

【0023】また、制御部113は、プログラマブル増幅器(PGA)201の増幅率を制御する。この場合、制御部113は、補正值テーブル204を用いることによって、音声認識部203で認識された各単語の音声信号レベルをその単語種毎に異なった補正值で補正し、その補正結果を利用してそのプログラマブル増幅器(PGA)201の増幅率を調整する。補正值テーブル204には、各単語種毎に異なった補正值が定義されている。このプログラマブル増幅器(PGA)201の増幅率の調整制御動作の詳細については、図3を参照して後述する。

【0024】音声辞書メモリ114には音声データが記憶されており、その音声データは規則合成音や音声フレーズから構成されている。規則合成音は、単音節音声に対応するものであり、また音声フレーズは単語や文等の複数の連続音節音声に対応するものである。この音声辞書メモリ114に格納されている音声データ(規則合成音または音声フレーズ)は制御部113によって選択指定され、その指定された音声データは音声辞書メモリ114から読み出されて音声出力部115に供給される。

【0025】音声出力部115は、音声辞書メモリ114から読み出された音声データを音声信号に変換し、その変換出力をネットワーク制御ユニット111に供給する。この音声出力部115は、例えば、音声データ受信バッファ、D-A変換器、および増幅器等によって構成されている。図2には、補正值テーブル204の具体的な内容の一例が示されている。

【0026】補正值テーブル204は、同一条件下(回線の減衰、発声の大小等)において単語種によって音声レベルの相違がなくなる様に入力音声信号レベルを補正するためのものであり、図示のように、音声認識部203で認識可能な全ての単語種、すなわち「イチ」、「ニ」、「サン」、「ハイ」、「イイエ」…についてそれぞれ補正值が定義されている。この補正值は、各単語種の音声信号レベルを最も発声レベルの高い単語種が入力された場合と同等の音声レベルに変換するための乗数である。

【0027】すなわち、単語「サン」、「ハイ」は発信者の発声レベルが最も高いので、これら単語「サン」、

5

「ハイ」の音声信号レベルを基準レベル（補正值＝“1”）とし、この基準値レベルと同等のレベルになるように他の単語に対する補正值が決められている。例えば、単語「イチ」の標準的な発声レベルは、最大で単語「サン」、「ハイ」の標準的な発声レベルの1/2であり、また単語「ニ」の標準的な発声レベルは、その単語「イチ」よりもやや高い程度であるので、ここでは、単語「イチ」に対する補正值として“2”、単語「ニ」に対する補正值として“1.9”が設定されている。さらに、単語「イイエ」の標準的な発声レベルは、単語

「ニ」よりも少し高い程度であるので、単語「イイエ」に対する補正值としては“1.5”が設定されている。次に、図3を参照して、制御部113によって実行されるプログラマブル増幅器（PGA）201の増幅率調整動作の原理を説明する。

【0028】ここでは、例として、入力音声信号の音声信号レベルが2[V_{P-P}]（ピークピーク値）で、認識部203によるその入力音声信号の認識結果が単語

「イチ」であった場合について考える。

【0029】この場合、補正值テーブル204からは単語「イチ」に対応する補正值“2”が読み出され、その補正值“2”と2[V_{P-P}]の入力音声信号レベルとの乗算が乗算器205によって行われる。この乗算結果である4[V_{P-P}]は、ゲイン算出回路206によって基準レベル10[V_{P-P}]と比較され、その比である2.5倍（＝10[V_{P-P}]/4[V_{P-P}]）がプログラマブル増幅器（PGA）201の増幅率として設定される。ここで、基準レベル10[V_{P-P}]の値はA/D変換器202で量子化できる最大値に設定されている。この結果、プログラマブル増幅器（PGA）201の増幅率は2.5倍に設定され、以降入力される音声信号に対して2.5倍増幅する。

【0030】ここでは、制御部113がハードウェアロジックによって増幅率を求める場合を説明したが、制御部113は、その制御部113内のマイクロプロセッサによってソフトウェア的にゲイン算出処理を実行することもできる。このゲイン算出処理については、図5のフローチャートを参照して後述する。次に、図4を参照して、発信者と音声応答装置11間でのメッセージの送受信動作を説明する。

【0031】まず、発信者が電話回線網13を介して音声応答装置11を呼出し、これによって音声応答装置11に対してサービス要求を行う。呼び出された音声応答装置11は、発信者に対してサービス内容を問い合わせるための音声データを音声辞書メモリ114から取り出し、それをサービスコード問い合わせ用の音声メッセージとして発信者に送信する。この後、発信者は、希望するサービスに対応したサービスコード（数値列）を音声によって音声応答装置11に送信する。

【0032】音声応答装置11は、発信者からのサービ

6

スコードによって発信者の要求するサービス内容を認識すると、そのサービスを提供しても良い発信者か否かを確かめるために、暗証番号問い合わせ用の音声データを音声辞書メモリ114から取り出し、それサービスコード問い合わせ用の音声メッセージとして送信する。この後、発信者は、所定の暗証番号（数値列）を音声によって音声応答装置11に送信する。

【0033】サービスを提供しても良い発信者であることが暗証番号によって確かめられると、音声応答装置11は、要求されたサービス内容に対応した音声データを音声辞書メモリ114から順次取り出し、サービスメッセージの送信を開始する。

【0034】このような一連の音声信号の授受において、音声応答装置11は、発信者から通知されたサービスコードや暗証番号を再確認するための音声メッセージの送信（認識した番号等の送信）も行う。この場合、発信者は、音声応答装置11から送信された確認のためのサービスコードや暗証番号が正しければ「ハイ」、正しければ「イイエ」、の音声を送信する。発信者から「イイエ」の音声を送信された場合は、音声応答装置11は、問い合わせ用メッセージを再度発信者に送信する。

【0035】発信者から送信されるサービスコード通知や暗証番号通知の受信処理においては、音声応答装置11では、プログラマブル増幅器（PGA）201の増幅率の調整が単語種毎に実行される。

【0036】例えば、サービスコードが「131」の場合には、発信者からは単語「イチ」、「サン」、「イチ」の順で音声信号が所定の間隔をおいて送信される。この場合、音声応答装置11の制御部113は、単語「イチ」、「サン」、「イチ」がそれぞれ認識される度に、図5のゲイン調整ルーチンを実行してプログラマブル増幅器（PGA）201の増幅率の調整を行う。

【0037】すなわち、図5のフローチャートに示されているように、制御部113は、まず、音声認識部203から認識単語を示すデータを受取り（ステップS11）、その認識単語（ここでは、「イチ」）に対応する補正值（“2”）を補正值テーブル204からリードする（ステップS12）。次いで、制御部113は、図示しないレベル検出器等で検出した単語「イチ」の入力音声レベル（2[V_{P-P}]）と補正值（“2”）の乗算を行い、比較レベル（4[V_{P-P}]）を求める（ステップS13）。

【0038】この後、制御部113は、その比較レベル（4[V_{P-P}]）と予め設定された基準レベル（10[V_{P-P}]）とを比較し、その比（2.5＝10[V_{P-P}]/4[V_{P-P}]）をプログラマブル増幅器（PGA）201の増幅率として求める（ステップS14）。そして、制御部113は、プログラマブル増幅器（PGA）201の増幅率が2.5倍に設定されるように制御

信号をそのプログラマブル増幅器 (PGA) 201 に供給する (ステップ S15)。

【0039】この結果、プログラマブル増幅器 (PGA) 201 の増幅率は、初期設定された所定値から 2.5 に変更される。この状態で、次の単語「サン」が入力されると、その単語「サン」の音声信号レベル (例えば、5 [V_{p-p}]) は、2.5 倍に増幅されて、A-D 変換器 202 に入力される。そして、音声認識部 203 によって単語「サン」が認識されると、前述したステップ S11~S15 のゲイン調整ルーチンが再び実行される。

【0040】この場合、単語「サン」の補正值は“1”なので、単語「サン」の音声信号レベル (5 [V_{p-p}]) はそのまま比較レベルとなり、10 [V_{p-p}] の基準レベルと比較される。この結果、プログラマブル増幅器 (PGA) 201 の増幅率は、2.5 から 2 に変更される。このようにして、単語を受け取る度にステップ S11~S15 の処理が繰り返し実行されて行く。

【0041】以上のように、この実施例においては、入力音声信号の各単語種は補正值テーブル 204 内の対応する補正值によって補正され、最も発声レベルの高い単語 (「サン」、「ハイ」) が入力された場合と同等の音声レベルに変換される。そして、その補正結果は所定の基準レベルと比較され、その比較結果に基づいてプログラマブル増幅器 (PGA) 201 の増幅率が制御される。

【0042】このため、各単語種毎に増幅率の調整度合いを変更できるようになるので、単語種による音声信号レベルの相違にとらわれず、常に入力した音声信号が最適なレベルになる様な増幅率の調整を行うことができ

る。

【0043】したがって、発声レベルの低い単語種 (例えば、「イチ」) の後に発声レベルの高い単語種 (例えば、「サン」) が入力された場合でも増幅率を上げ過ぎるといった不具合を防止でき、レベル過多等による誤認識を防ぐことができる。この結果、音声認識率が向上し、良好な音声応答サービスを提供できるようになる。

【0044】

【発明の効果】以上のように、この発明によれば、入力音声信号のレベル不足、又はレベル過多による誤認識が生じないように入力音声信号の増幅率の制御を適切に行えるようになり、十分に認識率の高い音声認識を行うことが可能となる。

【図面の簡単な説明】

【図1】この発明の一実施例に係る音声応答装置のシステム構成を示すブロック図。

【図2】同実施例の音声応答装置に設けられた補正值テーブルの具体的内容の一例を示す図。

【図3】同実施例の音声応答装置の入力ゲイン調節動作の原理を説明するためのブロック図。

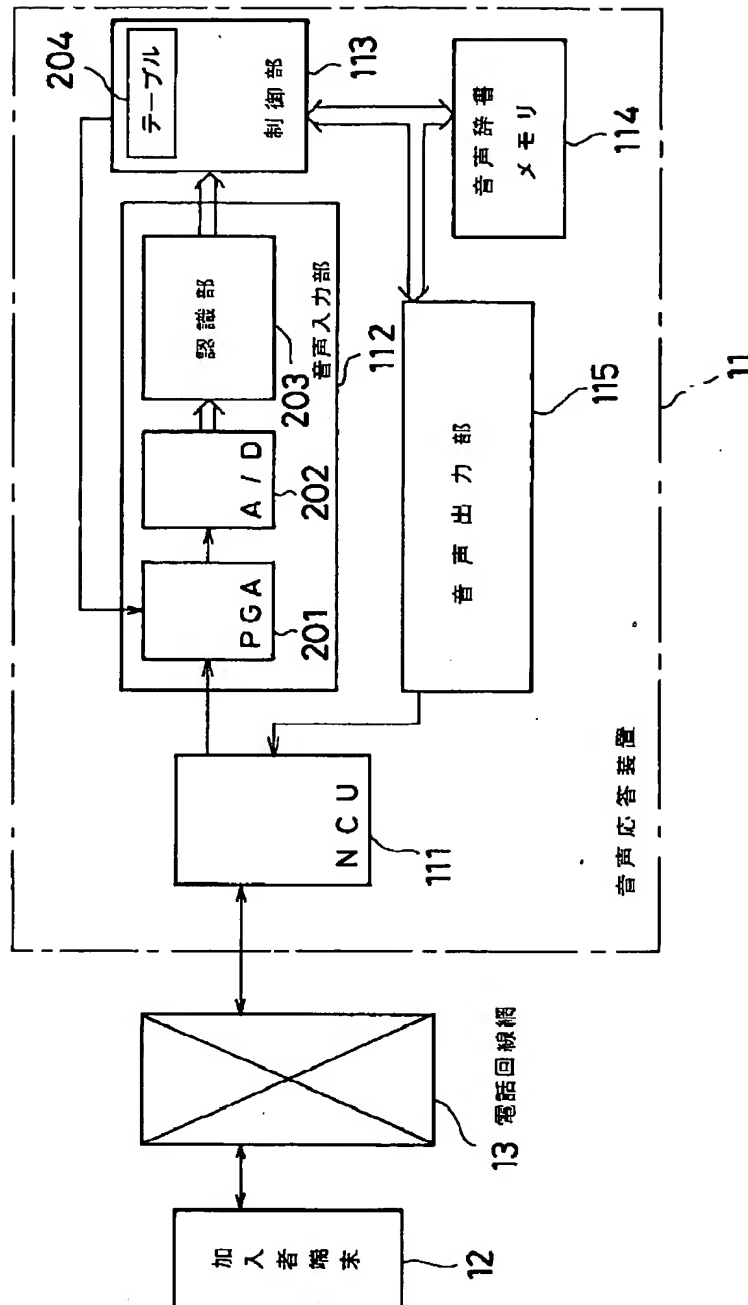
【図4】同実施例の音声応答装置と発信者との間における音声信号の発信動作を説明するための図。

【図5】同実施例の音声応答装置によって実行される入力ゲイン調節処理の動作手順を説明するフローチャート。

【符号の説明】

11…音声応答装置、12…加入者端末、13…電話回線網、112…音声入力部、113…制御部、201…プログラマブル増幅器 (PGA)、202…A-D変換器、203…音声認識部、204…補正值テーブル。

【図1】

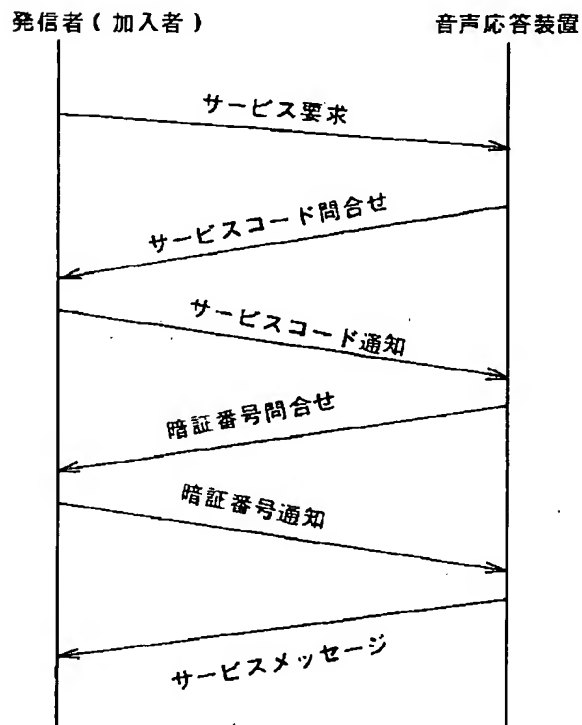


【図2】

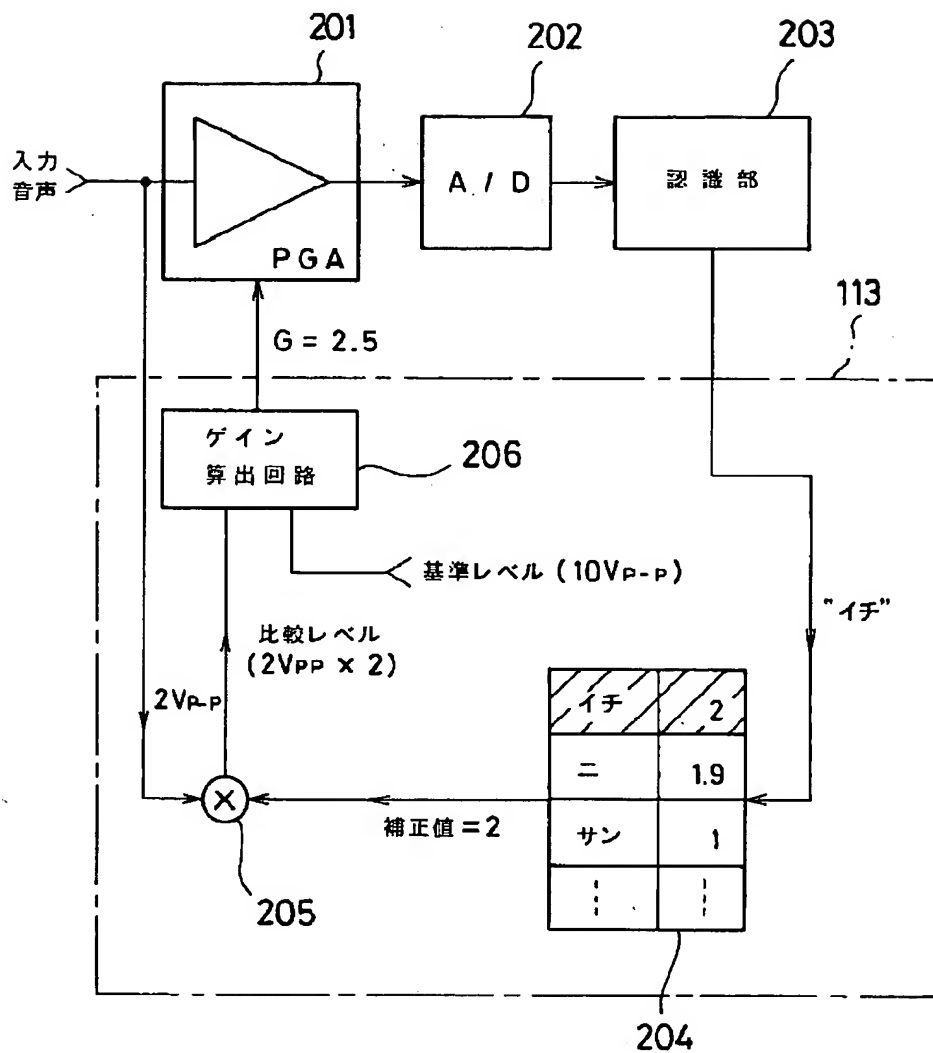
| 単語種 | 補正值 |
|--------|-----|
| イチ (1) | 2 |
| ニ (2) | 1.9 |
| サン (3) | 1 |
| ハイ | 1 |
| イイエ | 1.5 |
| ⋮ | ⋮ |

補正值テーブル
204

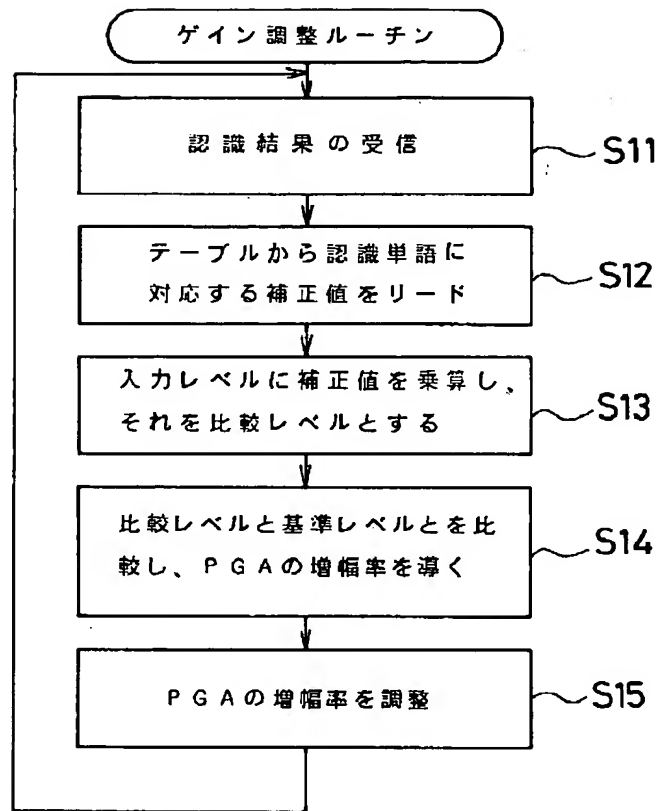
【図4】



【図3】



【図5】



DERWENT-ACC-NO: 1993-193901

DERWENT-WEEK: 199324

COPYRIGHT 1999 DERWENT INFORMATION LTD

TITLE: Voice response unit for telephone
network - has controller to adjust gain of voice
signal amplifier according to word voice signal level
NoAbstract

PATENT-ASSIGNEE: TOSHIBA KK[TOKE]

PRIORITY-DATA: 1991JP-0281540 (October 28, 1991)

PATENT-FAMILY:

| PUB-NO | PAGES | PUB-DATE | MAIN-IPC |
|---------------|-------------|--------------|----------|
| JP 05122366 A | | May 18, 1993 | |
| 009 | H04M 003/50 | | N/A |

APPLICATION-DATA:

| PUB-NO | APPL-DESCRIPTOR | APPL-NO |
|----------------|------------------|---------|
| JP 05122366A | N/A | |
| 1991JP-0281540 | October 28, 1991 | |

INT-CL (IPC): H04M003/50

ABSTRACTED-PUB-NO: JP 05122366A

EQUIVALENT-ABSTRACTS:

CHOSEN-DRAWING: Dwg.1/5

TITLE-TERMS: VOICE RESPOND UNIT TELEPHONE NETWORK CONTROL
ADJUST GAIN VOICE
SIGNAL AMPLIFY ACCORD WORD VOICE SIGNAL LEVEL
NOABSTRACT

DERWENT-CLASS: W01

EPI-CODES: W01-C02B4;

SECONDARY-ACC-NO:

Non-CPI Secondary Accession Numbers: N1993-148892

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINE(S) OR MARK(S) ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.